Phylogeographic structure in the CFR genus *Pauridia* revealed by inter-simple sequence repeat amplification

Marc Burman

BSc. Honours 2005

Systematics Project

Supervisor: Professor Terry Hedderson

University of Cape Town

Abstract

The Cape Floristic Region, South Africa, has high numbers of rare and endemic plants. Many, including *Pauridia* Harv., are geophytes in the Hypoxidaceae. The two species of *Pauridia* differ in morphology and range, with *P. minuta* occupying a wide range across lowland CFR and *P. longituba* being restricted to granite outcrops on the West coast. Genetic structure correlated to geography has been described for the haploid chloroplast genome. Here the diploid nuclear genome is investigated using inter-simple sequence repeat amplification. Eleven populations were sampled from DNA collected for a previous study. Analysis of molecular variance was done. Ordination based on similarity and covariance was done to detect structure. Most of the variance was found to be within populations (65.54%) and between populations within groups (35.69%). Some variance (10.10%) was found between *P. minuta* geographic groups. Principal Component Analysis revealed a little structure, grouping some similar haplotypes together. Cluster analysis placed *P. longituba* within *P. minuta*, reflecting the haplotypic structure. Differences in structure revealed between the chloroplast and nuclear genome may be explained by differences in gene flow resulting from the different modes of inheritance – chloroplast via seed and nuclear via pollen and seed. Further analysis of genetic distance correlations between the chloroplast and nuclear data would provide a useful quantitative measure of difference in structure.

Introduction

The Cape Floristic Region (CFR), although relatively small, has high numbers of rare and endemic plants (Cowling *et al.* 1996). It is a sclerophyllous Mediterranean vegetation type found only on nutrient-poor sandstone soils, which occurs only in the winter rainfall region of the Western Cape. The CFR is well known for its high species diversity, especially in terms of species turnover, and has been given high conservation status as a Global Biodiversity Hotspot (Myers *et al.* 2000). Species richness and the high levels of endemism in the Western Cape are thought to have arisen during a period of climatic stability that began in

the Pliocene, and were promoted by geological stability, steep ecological gradients and fine-scale habitat diversity (Goldblatt 1997).

The worlds greatest concentration of geophyte species (1552 species) is found in the CFR (Proches *et al.* 2005). *Pauridia* Harv. (Hypoxidaceae) is a geophytic genus endemic to the Cape Floristic Region (Rowe 2005). The genus consists of two species, *P. minuta* occurring throughout the Cape Floristic Region, on lowland areas of about 40 000 km^2, and *P. longituba* found only near Langebaan on the West coast (see **Figure 1**), The two species are morphologically distinct in the structure of their reproductive parts. In *P. longituba*, the perianth tube is extended, and the pedicel shortened, so that the ovary is entirely below ground. In *P. minuta* the ovary is on a longer pedicel, is above ground and the perianth tube is shorter. The genus belongs within the geophytic order Asparagales, in the family Hypoxidaceae. Together with the genus *Spiloxene*, it forms a clade containing of which 28 are CFR endemic species, out of 31 in total.

The species differ ecologically: *P. longituba* occurs only in sand-filled rocky crevices on granite boulders, while *P. minuta* occurs in such crevices, but also in sandstone crevices, on rock pavements and on clay soils. Population densities in rocky crevices can be extremely high, reaching tens to hundreds of thousands per m^2. The entire description of the genus given above is from the work of Rowe (2005).

Rowe (2005) has studied phylogeographic structure in *Pauridia* using chloroplast markers. Nested clade phylogenetic analysis (NCPA, Templeton 1998) as well as coalescent approaches were used to infer population structure, and to test hypotheses on the cause of this structure. Rowe (2005) revealed five geographically distinct two step clades for *P. minuta*: Southern Cape, West coast, Worcester, Langebaan and Paternoster. The genus consists of two species, *Pauridia minuta* and *P. longituba* (confirmed using Nested Clade Phylogenetic Analysis, Rowe (2005)). *Pauridia minuta* is found throughout the

range of the genus, except in populations around Paternoster and Saldanha, in which only *P. longituba* is found. Two populations in this area contain both species, although the *P. minuta* present in these populations is a geocarpic form, which is intermediate with *P. longituba*. One other population of the geocarpic form was found, at Langebaan. Rowe (2005) has proposed a number of processes by which the haplotype divergence between populations in these three areas could have occurred. These are that the mountains represent barriers to seed flow; or that during the last glacial period (circa 25000 years ago) the effect of climatic drying on the eastern CFR (which exposed a lowland corridor) caused the eastern *P. minuta* to retreat to mountain refugia, being unable to survive on the lowland corridor (Rowe 2005).

Two chloroplast regions, psbA-trnH, a highly variable intergenic spacer (Sang *et al.* 1997) and rps16, an intron of a ribosomal protein-encoding gene (Oxelman *et al.* 1997) revealed population level structuring, and clearly delineated the two species, *P. minuta* and *P. longituba* (Rowe 2005). The differing modes of inheritance of nuclear and chloroplast genes (chloroplast via seed, nuclear via seed and pollen) raises the potential for them to reveal different population histories. This study uses inter-simple sequence repeat (ISSR) amplification, first described by Zietkiewicz *et al.* (1994) to evaluate structure within the nuclear genome of the genus, and to compare this structure to that found for the chloroplast markers by Rowe (2005). Chloroplasts are maternally inherited, and their phylogeographic structure will thus reflect seed flow, while nuclear genes are transmitted by pollen and seeds, and so will reflect both mechanisms of dispersal.

Inter-simple sequence repeat amplification has been much used in plant genetic studies as a rapid genetic fingerprinting method to detect population structure (for example, Qian *et al.* 2001 in rice, Camacho & Liston 2001 in *Botrychium pumicola*, a fern). A number of studies have compared the use and effectiveness of unmapped markers such as ISSRs to that of mapped markers such as simple

sequence repeats (SSR), restricted fragment length polymorphisms (RFLP) and cleaved amplified polymorphic sequences (CAPS). For example, ISSRs and cleaved amplified polymorphic sequence (CAPS) were compared for *Arabidopsis thaliana* (Barth *et al.* 2002). Cluster analysis using similarity based on CAPS data grouped the populations geographically, while the ISSR data did not reflect any geographical population structure. It was argued that CAPS were better for diversity analyses in *Arabidopsis thaliana* as they are mapped and can thus be selected to represent the entire genome.

Inter-simple sequence repeats are DNA microsatellite markers, amplified using PCR, and visualized under UV light on an agarose gel. The process is rapid and highly reproducible (Yang *et al.* 1996 *in* Xiao *et al.* 2004), which is due to the high stringency of the primers used (Wolfe *et al.* 1998 *in* Xiao *et al.* 2004). ISSR markers have been shown to be hyper-variable and useful in population level studies (for example, Qian & Hong 2001; Xiao *et al.* 2004).

The aim of this study is to discover how phylogeographic patterns, as revealed by nuclear inter-simple sequence repeat analysis, differ from those revealed by analysis of chloroplast markers in the genus *Pauridia*. The hypothesis is that in structure as revealed by the two approaches should differ greatly. To test this I firstly measured the level of ISSR variation between populations and between the two species of *Pauridia*, and secondly used ordination techniques to detect population level structuring in the genus.

Methods

Samples were collected and extracted by Rowe (2005). Extractions were done according to the CTAB method (modified from Doyle and Doyle 1990; Rowe 2005). Annealing temperature (T_A) and Mg^{2+} concentration assays were run for both primers to determine the T_A and Mg^{2+} concentration that produced the clearest bands when run on agarose gel. Two primers were selected from an initial assay of 9, from the UBC ISSR primer set. Those tested were: 835, 841,

846, 812, 813, 851, 853, 858, 814. The two selected were 835 (GAGAGAGAGAGAGAGYC) and 841 (GAGAGAGAGAGAGAGAYC). The PCR master mixes were as follows based on the assay results: 13.35μl PCR water, 2.5μl of 10xBuffer, 3.0μl Mg^{2+}25mM, 1.0μl dNTPs 2.5mM, 1.0μl of 10uM Primer, 0.15μl Taq and 4μl template DNA. Nuclear DNA was PCR amplified on a GeneAmp® PCR system 2700, (Applied Biosystems) machine, on the following cycles: primer 835: 94°C for 1.30mins, 53°C for 1 min, 72°C for 1 min, 94°C for 30 seconds, 51°C for 2 minutes, 72°C for 3 minutes, with three holds; primer 841: 94°C for 1.30mins, 52°C for 1 min, 72°C for 1 min, 94°C for 30 seconds, 51°C for 2 minutes, 72°C for 3 minutes, with three holds, 4°C forever. PCR products were visualised using electrophoresis.

Negative controls (blanks) containing PCR water instead of DNA were run with products or all PCRs on 1% agarose check-gels to test for contamination. Only PCR runs with totally clear blanks were used for scoring.

Scoring gels were done using 1.5% agarose in 0.5% TBE solution. Three point five μl of Ethidium Bromide was added to the agarose-TBE solution. The loading ratio was 7.5μl of PCR product to 2.5μl of dye. One μl of Eco-Rv ladder was run at either end of the 17-well gel, and one in the middle of the gel. Gels were electrophoresed at 85V to the full length of the gel (about 2.5 hours). Gels were placed in a UVIDOC Uvitec Imaging Machine and photographed. Images were analysed using Arelequin v1.1 (Schneider *et al.* 1997) software package to detect bands, and visual inspection to verify bands and to score. Eleven populations were selected based on their representativeness of the geographic and hapltypic diversity of the genus. Five samples per population were used.

Statistical analysis of the scores was as follows: Analysis of Molecular Variance (AMOVA, Excoffier *et al.* 1992) was done using Arelequin v1.1 (Schneider *et al.*, 1997) software. Cluster analysis was done in NTSYS-pc (NTSYSpc, Numerical Taxonomy System, Version 2.2 for Windows, Rohlf 2000), using the unweighted

pairwise group means averaging (UPGMA) algorithm on Jaccard similarity values. Principal Component Analysis (PCA) in STATISTICA 7 was used to perform non-heirarchical ordination of the populations. Bands were scored as present (1) or absent (0) for each accession. Band presence was converted to band frequency in each population, and this was used in the cluster analysis and ordination. I tested for difference in the contribution of bands to factors in the PCA using a non-parametric test, Friedman ANOVA and Kendall Coefficient of Concordance.

Results

Band Frequency

No bands were restricted to one population. Of the 11 populations and 25 bands scored, 2 bands were absent from 1 population and 7 bands were absent from 2 populations, with none absent from more than 2 populations. One population (Napier) had all bands polymorphic (**Figure 2**). Three populations (Bonteheuwel , Witteklip Farm M and Paternoster) had no bands totally absent (**Table 3**, **Figure 2**). Three populations had high levels of polymorphism (80%: Simonstown, Rondeberg and Bonteheuwel), 2 more had greater than 50% polymorphic bands, and the rest had more than 48% bands present (**Table 3**).

Analysis of Molecular Variance

The vast majority of molecular variation is found between populations within groups (35.69%) and within populations (65.54%) (**Table 5**). No significant variation occurred between P. minuta and P. longituba.

There was significant variation (10.10%) among the P. minuta geographic groups (South Coast – Heidelberg, Napier and Botrivier; West Coast – Rondeberg, Langebaan, Bonteheuwel and Witteklip Farm; Simonstown and Worcester, **Table 6**). Among populations within groups 21.88% variation occurred, and 68.02% within populations (**Table 5**). No significant variation occurred between *P. minuta* and *P. minuta* geocarpic, while 31.12% and 70.36% of the variation occurred

among populations within groups, and within populations respectively. The three *P. longituba* populations varied widely and significantly ($p<0.001$, **Table 5**). The most divergent were Witteklip Farm *P. longituba* and Kleinberg with FST=0.622, and the least were Witteklip Farm and Paternoster with FST= 0.477 (**Table 5**).

Principal Component Analysis (PCA)

The ordination of the populations in the PCA of all populations (**Figure 4**) reflects some concordance with the cladistic analysis of Rowe (2005; **Table 4** and **Figure 3**, reprinted with permission). The projection of the cases on the 2nd and 3rd factors (explaining 16.80% and 15.07% of the variance respectively, **Figure 4**) produced a pattern in which closely related haplotypes were placed near to eachother. I selected the 2nd and 3rd factors for presentation in the results, as the populations appeared to be grouped on them in a more meaningful way than on the first two factors. I have included the projection on the first two factors as an appendix. Group 1 (**G-1**) contains haplotypes G,H,I,J (all within 1 step) and O, which is five steps from I. Haplotypes H (*P. minuta*, Simonstown) and J,G and I (*P. minuta*, Rondeberg) are within 1 step, and were near eachother in the PCA (**Figure 3**). Haplotype I is found in six populations, three *P. minuta* and three *P. longituba* (**Table 4** and **Figure 3**, Rowe 2005), making it the most common haplotype (Rowe 2005). Witteklip Farm L (*P. longituba*, haplotype I) is near Rondeberg and Simonstown in **Figure 2**. Group 2 (**G-2**) contains haplotype P (Napier and Heidelberg), which is unique to the south coast *P. minuta* populations, and is four steps from I, and one step from O. Groups 3 (**G-3**) contains haplotypes A,B and C, which are on the same branch of the haplotype network in Rowe (2005, **Figure 3**). Haplotype C (*P. minuta*, Langebaan) is four steps from haplotype I (**Figure 3**, Rowe 2005). Haplotype B (*P. minuta*, Bonteheuwel) is one step from C, and A (*P. longituba,* Paternoster), is 11 steps from haplotype C (*P. minuta*, Langebaan). Haplotype F (*P. longituba*, Kleinberg) is one step from haplotype I.

Cluster Analysis

Cluster analysis (**Figure 5**) resulted in a tree that grouped populations together which have highly divergent haplotypes. Only Heidelberg and Botrivier (haplotypes P and O) were grouped together in agreement with the haplotype network (**Figure 3**, Rowe 2005). However, some of the groupings reflect geographical proximity (Heidelberg and Botrivier; Witteklip Farm *P. minuta*, Paternoster and Kleinberg, **Figure 5**). Overall levels of similarity were less than 40%.

Discussion

The main questions in this study were, "What genetic structure does inter-simple sequence repeat data reveal, and how does this compare to the structure revealed by chloroplast marker sequences?" I have used a statistical approach and two graphical tools to analyse the ISSR data. The statistical approach (analysis of molecular variance) revealed significant variation between populations, but not between groups of populations except in one instance (between the four geographical groups of *P. minuta*). The variation among populations within groups was 35.69% (among the groups *P. minuta* and *P. longituba*), 31.12% (between the groups *P. minuta* and *P. minuta* geocarpic form), and 21.88% (in the *P. minuta* geographic groups) (**Table 5**), which is similar to the average for RAPD markers, 35%, as ascertained from a metanalysis of 108 RAPD-based studies (Nybom & Bartish 2000 *in* Nybom 2004).

The levels of within population diversity are high (65.54% in the *P. minuta* and *P. longituba* comparison, 68.02% in the *P. minuta* geographic groups comparison and 70.36% in the comparison of *P. minuta* and its geocarpic form, **Table 5**). Such high levels are not unusual – ISSR generally produces high within-population level variation (Nybom 2004), and attributes more variation to lower levels (within populations) than higher levels (between populations or groups) (Quian *et al.* 2001). In a study of ploidy levels and geographic origins in *Buchloe*

dacytloides in which nuclear and cytoplasmic markers were compared, nuclear markers (which included ISSR) were found to be the more informative, detecting more variation than cytoplasmic markers (Budack *et al.* 2005). The levels of variation found here lend credibility to the data, as they are consistent with those found in other studies (eg. Quian *et al.* 2001).

There was no variation between the two species (**Table 5**). This result reflects the haplotype network structure in Rowe (2005) in which the *P. longituba* haplotypes are dispersed amongst the *P. minuta* haplotypes (**Figure 3**). The 10.10% variance found between the *P. minuta* geographic groups also reflects the haplotype network structure, in which the geographic groups are widely separated (**Figure 3**). The high variance found between the *P. longituba* populations (**Table 5**) reflects the separation of the populations on the haplotype network (**Figure 3**).

The grouping of the populations in the principal component analysis had a similar structure to the haplotype network by Rowe (2005) (**Figure 4**). The cluster analysis (**Figure 5**) showed less concordance with the haplotype network The similarity between Napier and Heidelberg (both haplotype P) is reflected here, while in the cluster analysis these populations are not related (**Figure 5**). The closely related haplotypes I,H,J and G are grouped in the upper left of the PCA, and the more distant M,N and K,L and the group of A,B and C are separated away from the top left. The only anomalies are Kleinberg (F) and Botrivier (O). Kleinberg is far from I in the PCA, while it is one haplotype from I in the haplotype network. Botrivier is four haplotypes from haplotype I in the haplotype network (Rowe 2005) but groups with it in the PCA, rather than with Napier and Heidelberg, to whom it is closer geographically and in haplotype. These two anomalies are further complicated in the cluster analysis (**Figure 5**). Firstly, Botrivier is placed closest to Heidelberg, while Napier is most distant from Heidelberg; secondly, Kleinberg is placed closest to Paternoster and Witteklip Farm (haplotype MN), which all are closely geographically related (see **Figure 1**).

So, while Botrivier was grouped with geographically closer populations by the PCA, in the cluster analysis it was nearest to Heidelberg

The differences between the haplotype structure (Rowe 2005), which is strongly geographical, and the ISSR patterns, might be attributable to the differences between seed and pollen flow. The results support the hypothesis that there should be general agreement between the two methods, with some differences. The analysis of molecular variance revealed variance that corresponded with the haplotype network of Rowe (2005). The two ordination methods used here resulted in different patterns, and only the principal component analysis agreed at all with the haplotype network structure of Rowe (2005). In the PCA, one discrepancy was in the distance of Kleinberg from the I haplotype. This is repeated in the cluster analysis (**Figure 5**), in which Kleinberg is far removed from the I haplotype. This suggests that the Kleinberg population has had pollen transfer with its neighbouring populations at Witteklip Farm and Paternoster, without receiving their chloroplast haplotypes at (haplotypes MN and A). This implies that seed flow has not recently taken place into Kleinberg. The isolation of Worcester from all other groups in the PCA mirrors its isolation in the haplotype network of Rowe (2005), and the same is true for Witteklip Farm *P. minuta*. The lack of agreement between the two ordination methods might be due to their different treatments of the data - while in this case the PCA was done as graphical display of covariance, the cluster analysis was based on similarity.

To summarise the results: firstly, the ISSR approach has revealed hypervariability within populations. Secondly, some geographic structuring of *P. minuta* has been shown, with 10% variance between the South coast, West coast, peninsula and Worcester groups. Thirdly, the two species were not separated, nor were the geocarpic and non-geocarpic forms of *P. minuta*. Lastly, the structure revealed by the ordination of ISSR data was partly similar to that in the haplotype analysis of Rowe (2005).

It would have been useful in this analysis to directly compare genetic distances between populations produced by each method (chloroplast sequencing and ISSR), by plotting the distances from one approach against the other. The correlation between them would describe quantitatively how differently seed and pollen flow create structure. The significance of the correlation would be very useful in making inferences. In this case such an analysis was not possible as a distance matrix was not produced in the original study by Rowe (2005), and there was not time during this study to produce one for the chloroplast sequence data of Rowe (2005).

References

Barth, S. Melchinger, A.E. & Lubberstedt, T.H. 2002. Genetic diversity in *Arabidopsis thaliana* L. Heynh. Investigated by cleaved amplified polymorphic sequence (CAPS) and inter-simple sequence repeat (ISSR) markers. *Molecular Ecology* **11**: 495-505.

Budack, H., Shearman, R.C., Gulsen, O & I. Dweikat. 2005. Understanding ploidy complex and geographic origin of the *Buchloe dactyloides* genome using cytoplasmic and nuclear marker systems. *Theoretical and Applied Genetics* .

Cowling, R.M., Rundel, P., Lamont, B., Kalin, M., Arroyo, M. & M. Arianoutsou. Plant diversity in mediterranean-climate regions. *Trends in Ecology and Evolution*. **11**(9), 362-366, 1996.

Cowling, R.M. (ed.) 1992. The Ecology of Fynbos: Nutrients, Fire and Diversity. Oxford University Press, Cape Town.

Doyle,J. J. & J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15.

Excoffier, L., Smouse, P. & Quattro, J. 1992 Analysis of molecular variance inferred for molecular distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Goldblatt, P. 1997. Floristic diversity in the Cape Flora of South Africa. *Biodiversity and Conservation* **6**(3): 359-377.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B & J. Kent. 2002. Biodiversity hotspots for conservation priorities. *Nature* **403**: 853 – 858.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* **13**: 1143-1155.

Oxelman, B. Linden, M. & D. Berglund. 1997. Chloroplast rps16 intron phylogeny of the tribe *Sileneae* (*Caryophyllaceae*). *Plant Systematics and Evolution* **206**: 393-410.

Proches, H., Cowling, R.M. & D. du Preez. 2005. Patterns of geophyte diversity and storage organ size in the winter-rainfall region of southern Africa. *Diversity and Distributions* **11**: 101–109.

Qian, W. Ge, S. & D-Y. Hong. 2001. Genetic variation within and among populations of a wild rice *Oryza granulata* from China detected by RAPD and ISSR markers. *Theoretical and Applied Genetics* **102**: 440-449.

Rohlf, J.F. 2000. NTSYS-pc: numerical taxonomy and multivariate analysis system. Exeter Software, Setauket, New York.

Rowe, G. 2005. Speciation and phylogeography: Coalescent-based models applied to the Cape plant genus *Pauridia*. MSc. Thesis submitted October 2005, University of Cape Town.

Sang, T., Crawford, D.J. & T.F. Stuessy. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (*Paeoniaceae*). *American Journal of Botany* **84**: 1120-1136.

Schneider, S., Kueffer, J.-M., Roessli, D. & L. Excoffier. 1997. Arlequin ver 1.1. Software Package. http://anthropologie.unige.ch/arlequin.

Templeton, A.R. 1998. Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7**: 381-397.

Xiao, L-Q., Ge, X-J., Gong, X., Gang, H. & S-X. Zheng. 2004. ISSR variation in the endemic and endangered plant *Cycas guizhouensis* (Cycadaceae). *Annals of Botany* **94**: 133-138.

Zietkiewicz, E., Rafalski, A. & Labuda, D. 1994. Genome fingerprinting by simple sequence repeat (SSR) – anchored polymerase chain reaction amplification. *Genomics* **20**: 176-183.

Acknowledgements

I am very grateful to Graham Rowe for the use of his *Pauridia* DNA extractions, allowing me to reproduce his figures and tables, and for his enthusiasm. Thanks to Professor Terry Hedderson for help with the project and for funds and materials. Many thanks also to Julia Watson for endless support and encouragement.

Tables and Figures

Table 1. List of *Pauridia* populations sampled. Reproduced from Rowe (2005) with permission.

Number	Name	Co-ordinates	Herbarium sheet
1	Heidelberg	34 02 S 20 56 E	G. Rowe 1
2	Swellendam	34 04 S 20 27 E	G. Rowe 2
3	De Hoop	34 27 S 21 25 E	G. Rowe 3
4	Napier	34 28 S 19 55 E	G. Rowe 4
5	Gaansbaai	34 35 S 19 29 E	G. Rowe 5
6	Greyton	34 05 S 19 32 E	G. Rowe 6
7	Botrivier	34 22 S 19 07 E	G. Rowe 7
8	Worcester	33 37 S 19 22 E	G. Rowe 8
9	Simonstown	34 11 S 18 23 E	G. Rowe 9
10	Rondebosch	33 57 S 18 29 E	G. Rowe 10
11	Somerset West	32 50 S 18 03 E	G. Rowe 11
12	Paarl	33 45 S 18 58 E	G. Rowe 12
13	Allesverloren	33 24 S 18 51 E	G. Rowe 13
14	Rondeberg	33 25 S 18 16 E	G. Rowe 14
15	Langebaan	33 06 S 18 03 E	G. Rowe 15
16	Bonteheuwel	33 03 S 18 03 E	G. Rowe 16
17*	Witteklip Farm	32 56 S 17 58 E	G. Rowe 17
18*	Witteklip Beacon	32 55 S 17 58 E	G. Rowe 18
19	Saldanha	32 59 S 17 56 E	G. Rowe 19
20	Kleinberg	32 53 S 18 07 E	G. Rowe 20
21	Patrysberg	32 50 S 18 03 E	G. Rowe 21
22	Saint Helena	32 46 S 18 02 E	G. Rowe 22
23	Paternoster	32 48 S 17 55 E	G. Rowe 23

Table 2. List of Rowe's (2005) populations subsampled in this study. Numbers agree with those in **Figure 1**.

Number	Name
1	Heidelberg
4	Napier
7	Botrivier
8	Worcester
9	Simonstown
14	Rondeberg
15	Langebaan
16	Bonteheuwel
17	Witteklip Farm
20	Kleinberg
23	Paternoster

Table 3. List of populations, and frequency of absent, polymorphic, and present bands for all populations. f=frequency, abs=absent, poly=polymorphic, pres=present.

	f(abs)	f(polym)	f(pres)
Langebaan	0.16	0.24	0.60
Heidelberg	0.12	0.28	0.60
Botrivier	0.12	0.36	0.52
Worcester	0.12	0.36	0.52
Kleinberg L	0.16	0.36	0.48
Witteklip Farm L	0.08	0.44	0.48
Paternoster L	0.00	0.56	0.44
Witteklip Farm M	0.00	0.60	0.40
Simonstown	0.04	0.80	0.16
Rondeberg	0.04	0.80	0.16
Bonteheuwel	0.00	0.80	0.20
Napier	0.00	1.00	0.00

Table 4. Sequencing results for *Pauridia minuta* and *Pauridia longituba*, giving the haplotypes found for each individual. Reproduced from Rowe (2005) with permission.

Population #	Number of individuals	Haplotype frequency	Random sample
1	5	5P	P
2	5	5Q	Q
3*	4	3P, 1R	P
4	3	3P	P
5	5	5P	P
6	3	3P	P
7	4	4O	O
8*	5	4K, 1L	K
9	5	5H	H
10	5	5J	J
11	4	4J	J
12	4	4I	I
13	5	5I	I
14*	10	5J, 4G, 1I	G
15	9	9C	C
16	6	6B	B
17*	4	3M 1N	M
18	5	5M	M
Total	92 individuals	14 haplotypes	11 haplotypes

Table 3.1: *Pauridia minuta* sequencing results. For each population the number of individuals and the haplotype frequency is given. See section 3.4 'metapopulation structure' for an explanation of the column 'Random sample'. Only populations marked with a * contain more than one haplotype.

Population #	Number of individuals	Haplotype frequency	Random sample
17	5	5I	I
18	10	10I	I
19	2	2I	I
20	5	5F	F
21	5	5I	I
22*	3	2E 1D	E
23	3	3A	A
Total	33 individuals	5 haplotypes	4 haplotypes

Table 3.2: *Pauridia longituba* sequencing results. For each population the number of individuals and the haplotype frequency is given. See section 3.4 'metapopulation structure' for an explanation of the column 'Random sample'. Only populations marked with a * contain more than one haplotype.

AMOVA RESULTS

Table 5. Analysis of molecular variance of *Pauridia* populations. Significant percentages variation are in bold.

P. minuta versus *P. longituba*:

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	4.592	-0.01560 Va	-1.23
Among populations within groups	10	49.844	0.45243 Vb	**35.69**
Within populations	100	83.072	0.83072 Vc	**65.54**
Total	111	137.509	1.26755	

P. minuta geographic groups (see **Table 6**)

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	3	18.335	0.13484 Va	**10.10**
Among populations Within groups	5	17.628	0.29208 Vb	**21.88**
Within populations	75	68.097	0.90796 Vc	**68.02**
Total	83	104.060	1.33489	

Table 5 (continued). Analysis of molecular variance of *Pauridia* populations. Significant percentages variation are in bold.

P. minuta and *P. minuta* geocarpic:

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	5.399	-0.01902 Va	-1.47
Among populations within groups	7	30.563	0.40155 Vb	**31.12**
Within populations	75	68.097	0.90796 Vc	**70.36**
Total	83	104.060	1.29049	

Table 6. Geographical *P. minuta* groups used in the analysis of molecular variance (**Table 4**).

South coast	Peninsula	West coast	Worcester
Heidelberg	Simonstown	Rondeberg	Worcester
Napier		Langebaan	
Botrivier		Bonteheuwel	
		Witteklip Farm	

Table 7. Analysis of variance within *P. longituba*. FST distances are below the diagonal, significance p-values are above. Distance method: Pairwise difference. Wit: Witteklip Farm *P. longituba*; Kln: Kleinberg; Pnr: Paternoster.

	Wit	Kln	Pnr
Wit		**0.0059**	**0.0158**
Kln	0.622		**0.0099**
Pnf	0.477	0.492	

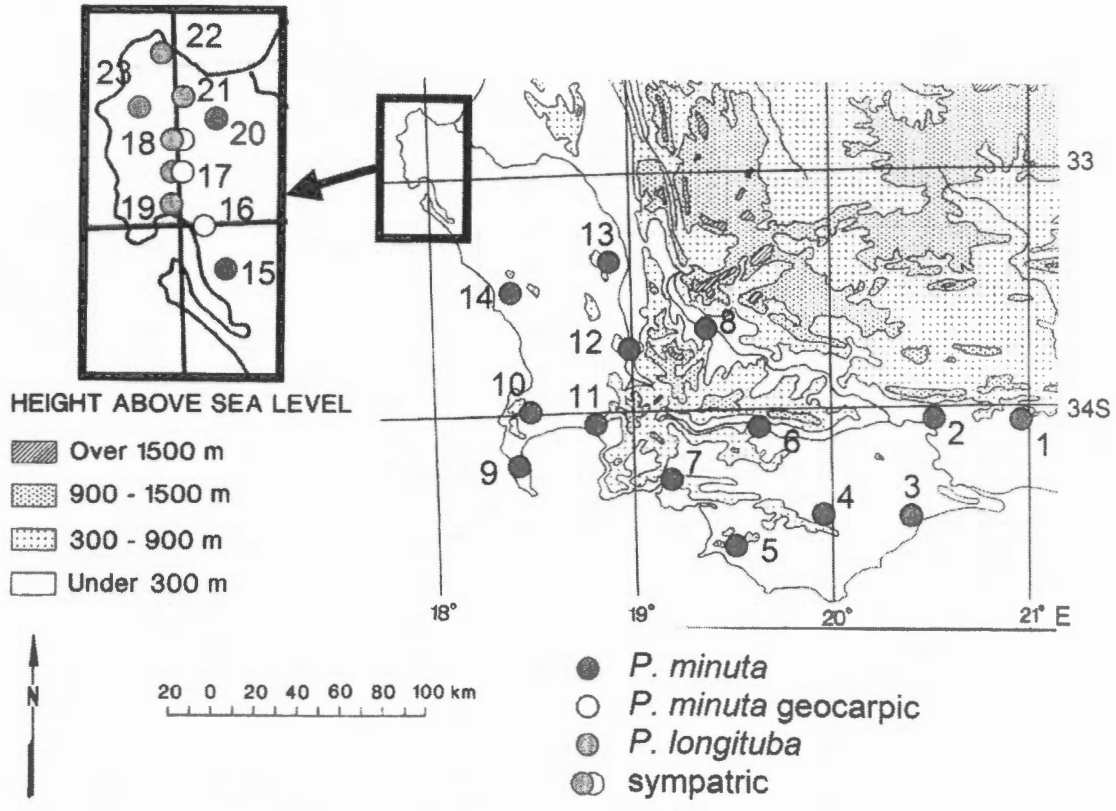

Figure 1. Map of the Western Cape, South Africa, showing *Pauridia* sampling locations. Reproduced from Rowe (2005) with permission.

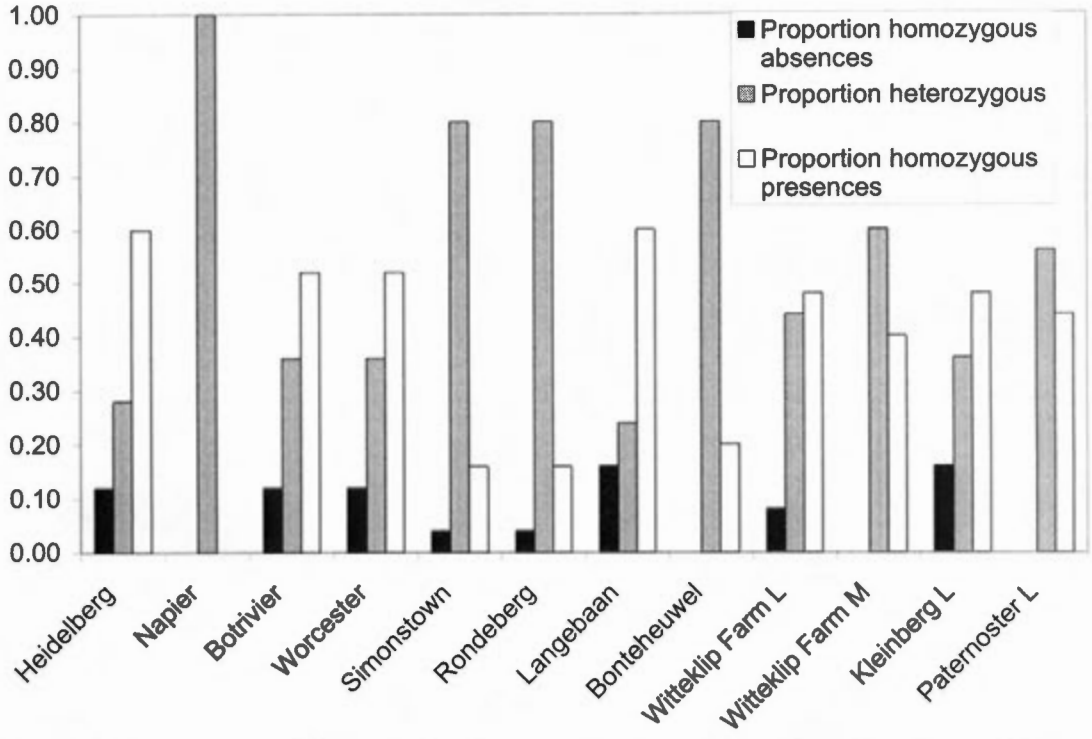

Figure 2. Frequency of band (allele) mono- and polymorphism in all populations.

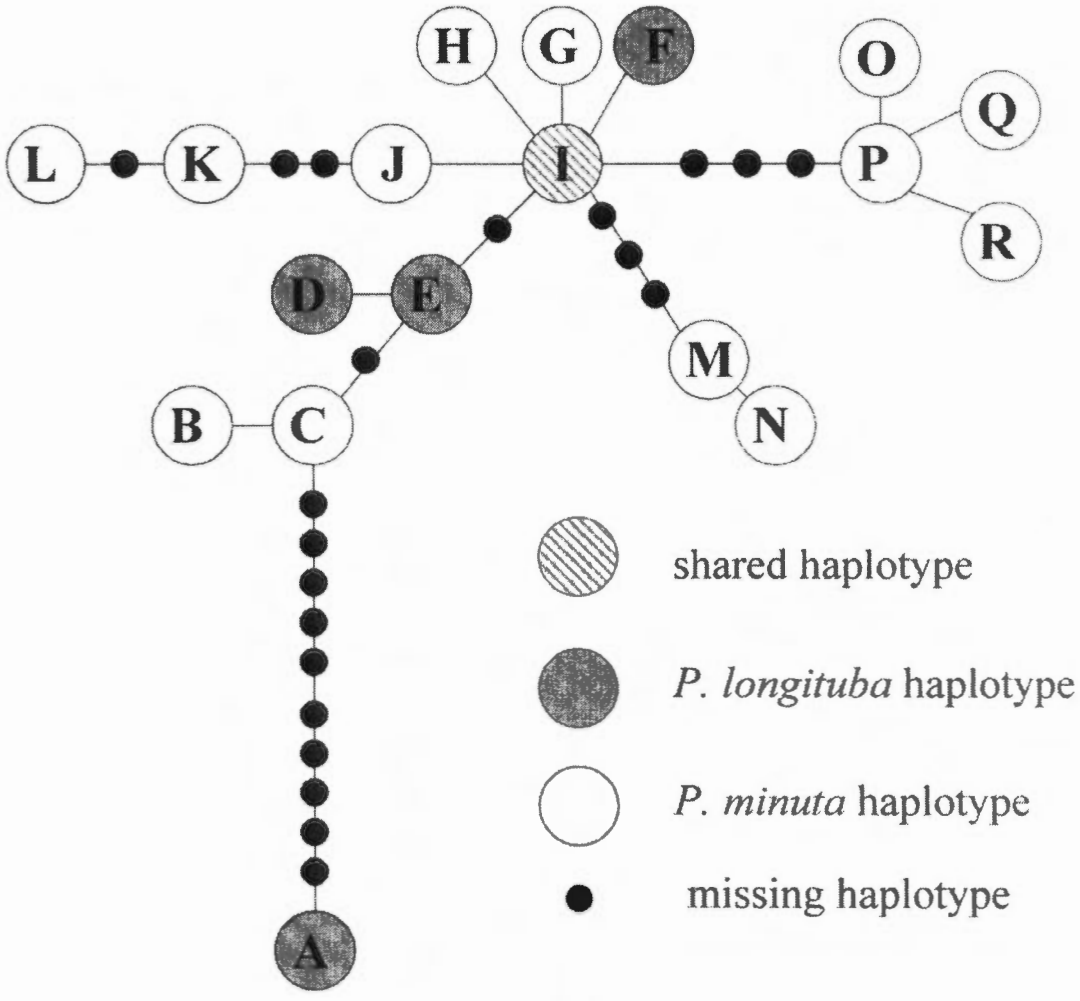

Figure 3.1. Pauridia haplotype network. Nodes on the network are haplotypes These either belong to one species, or both, or were not found in the sample (missing) and are inferred. Lines connecting haplotypes indicate a single mutational event on the 1188bp sequence.

Figure 3. *Pauridia h*aplotype network. Reproduced from Rowe (2005) with permission.

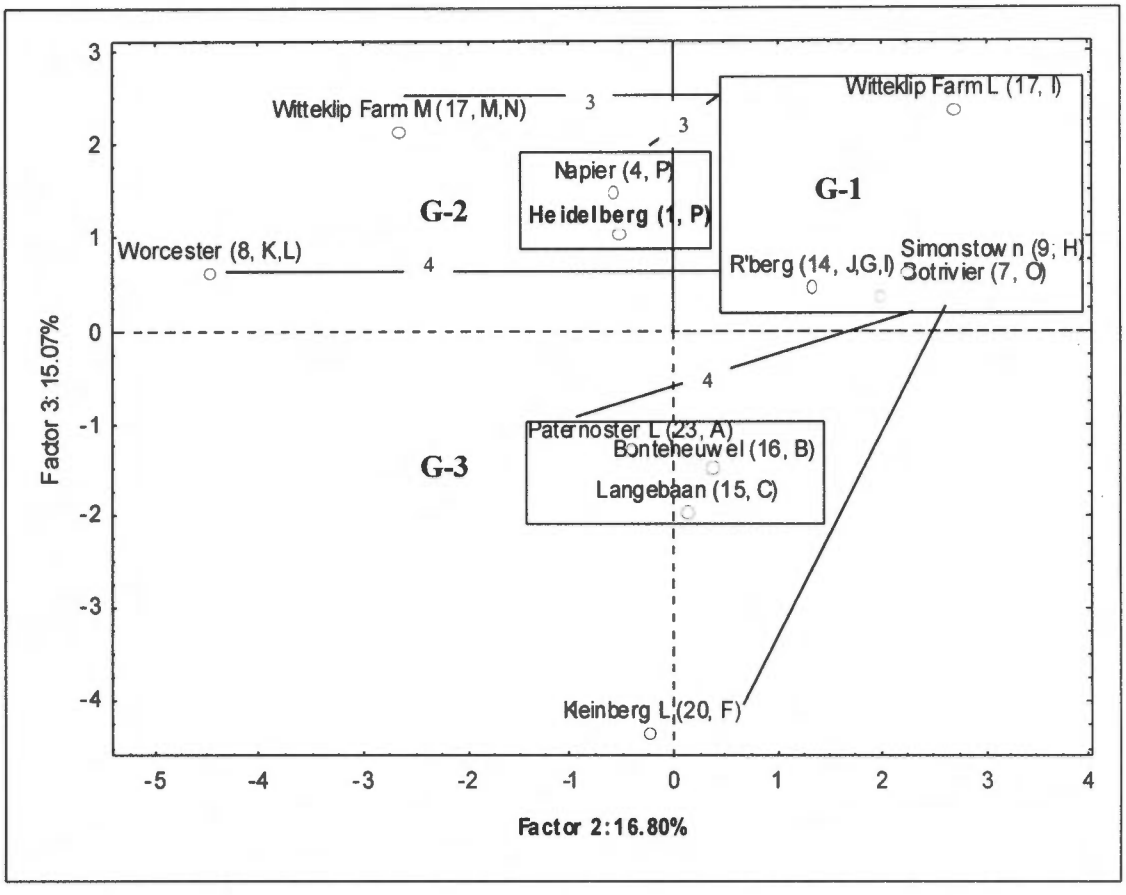

Figure 4. Principal component analysis of *Pauridia* populations, factor 2 and 3. In brackets after population names are population numbers and haplotype letters (haplotypes of Rowe 2005). Numbers on lines are average numbers of haplotypes between groups or populations. The M and L after Witteklip Farm denote sympatric populations of the two species, *P. minuta* and *P. longituba*.

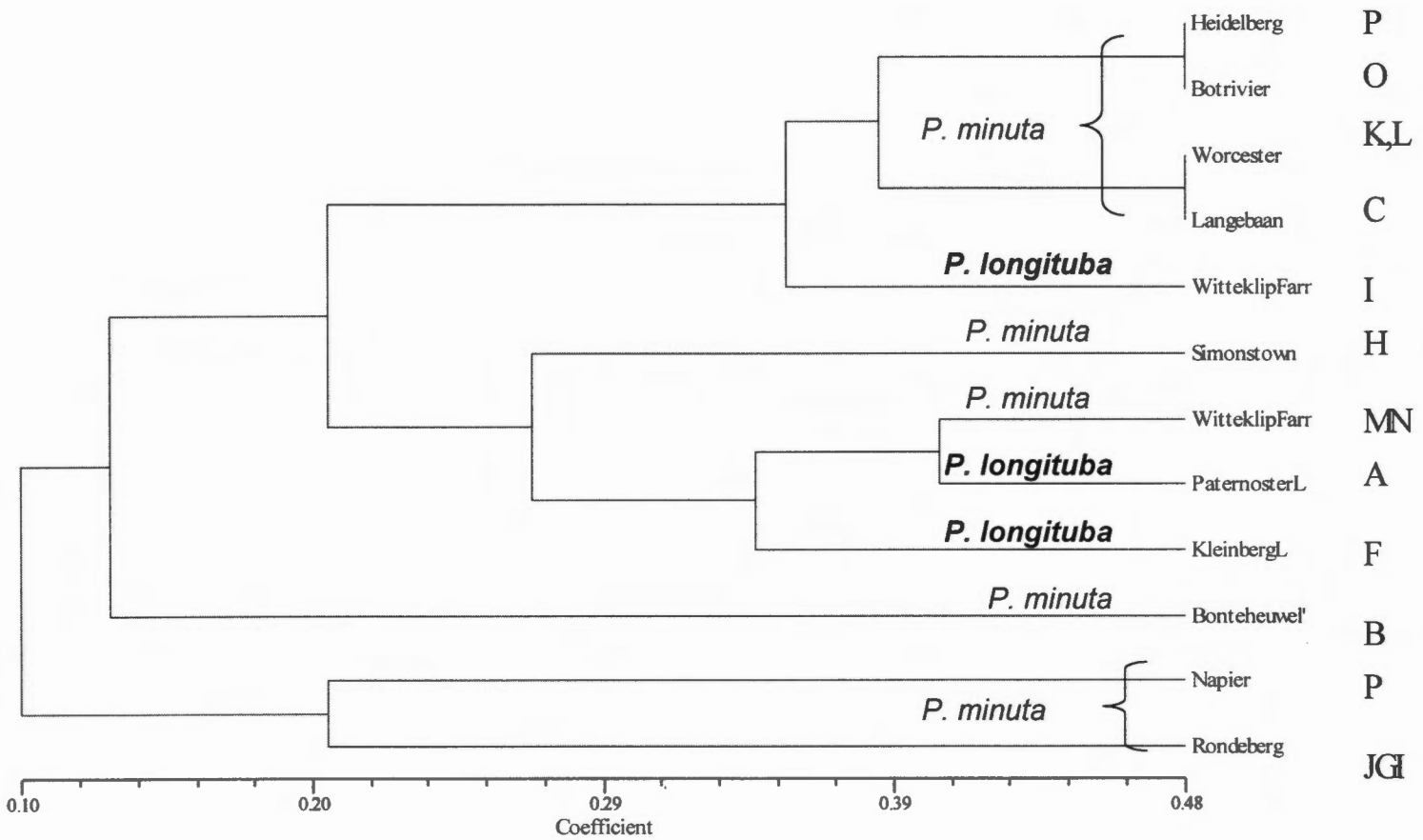

Figure 5. UPGMA cluster analysis dendrongram based on Jaccard similarity coefficients, of all populations, with population names. Haplotype letters (from Rowe 2005) are enlarged. The coefficient is % similarity.

Appendix

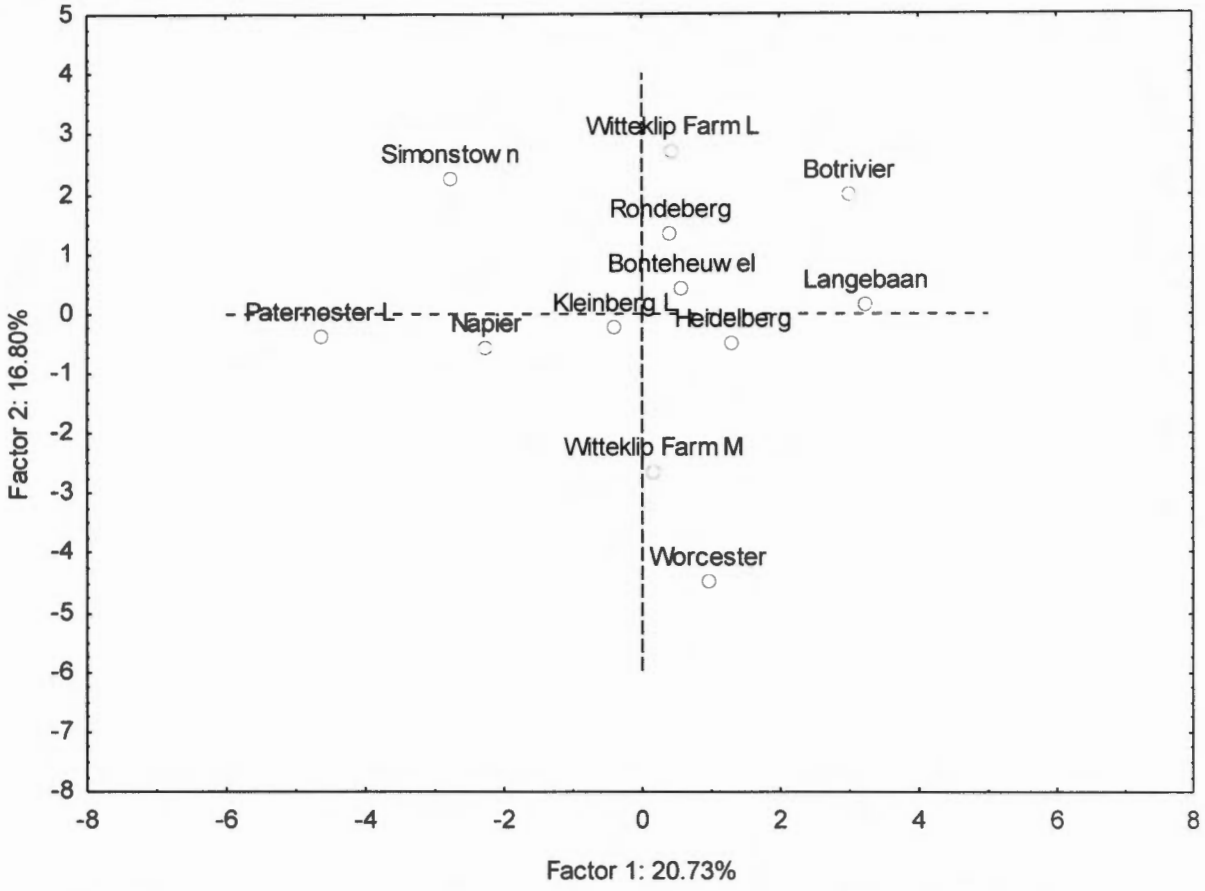

Figure 6. Projection of the populations on the first two factors in the principal component analysis of all the *Pauridia* populations sampled.